Часть 1. Тест.

Вопрос 1 ♣

Рассмотрим модель $Y_i = \mu + \varepsilon_i, E(\varepsilon_i) = 0, cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j, Var(\varepsilon_i) = \sigma_\varepsilon^2/X_i, i = 1, \ldots, 4$ при $X = (1, 2, 3, 4)$, оцененную обычным и обобщенным МНК. Во сколько раз дисперсия оценки коэффициента μ для модели, оцененной обобщенным МНК с учетом особенностей ковариационной матрицы ошибок, будет меньше дисперсии оценки, полученной обычным МНК?

A 10

B 30

C 50

D 100

E $\sqrt{10}$

F *Нет верного ответа.*

Вопрос 2 ♣

Для модели $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ известно, что $cor(X_1, X_3) = cor(X_2, X_3) = 0, cor(X_1, X_2) = r$, где $r = -0.9$. Параметр обусловленности для матрицы $(\widetilde{X}^T \widetilde{X})$, где $\widetilde{X}$ - матрица центрированных и нормированных значений регрессоров, равен

A $\sqrt{(1-r)/(1+r)}$

B $\sqrt{(1-r\sqrt{2})/(1+r\sqrt{2})}$

C $\sqrt{(1+r^2)/(1-r^2)}$

D $\sqrt{(1+r\sqrt{2})/(1-r\sqrt{2})}$

E $\sqrt{(1+r)/(1-r)}$

F *Нет верного ответа.*

Вопрос 3 ♣

При проверке гипотезы $H_0 : g(\beta) = 0$ для параметров модели $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon, \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$ с помощью теста Вальда, необходимо знать оценки параметров

A Регрессии на константу

B Регрессии на все факторы, кроме константы

C Только модели без ограничений

D Только модели с ограничениями

E Как модели с ограничениями, так и модели без ограничений

F *Нет верного ответа.*

Вопрос 4 ♣

Предельный эффект для непрерывной j-ой объясняющей переменной в логит модели рассчитывается по формуле

A $\hat{\beta}_j \frac{\exp(-Z)}{(1+\exp(-Z))^2}$

B $\hat{\beta}_j \frac{1}{(1+\exp(-Z))^2}$

C $\hat{\beta}_j \frac{1}{(1+\exp(Z))^2}$

D $\hat{\beta}_j \frac{1}{\sqrt{2\pi}\exp(-\frac{Z^2}{2})}$

E $\hat{\beta}_j \frac{\exp(-Z)}{(1+\exp(Z))^2}$

F *Нет верного ответа.*

Вопрос 5 ♣
Если площадь под ROC кривой для модели бинарного выбора равна 0.5, то качество предсказания модели

- [A] Не определяется этим показателем
- [B] Аналогично подбрасыванию симметричной монетки
- [C] Лучше, чем подбрасывание симметричной монетки
- [D] Хуже, чем подбрасывание симметричной монетки
- [E] Максимально возможное
- [F] *Нет верного ответа.*

Вопрос 6 ♣
Выборочная корреляция между переменными X_1 и X_2 равна 0.5. VIF для переменной X_1 в регрессии $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon, k > 4$

- [A] Не менее 4/3
- [B] Не более 4/3
- [C] Не менее 2
- [D] Не более 0.5
- [E] Не более 2
- [F] *Нет верного ответа.*

Вопрос 7 ♣
Для получения оценок LASSO регрессии $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$ используется критерий

- [A] $\min\limits_{\hat\beta}(RSS + \lambda \sum_{j=0}^{k} \hat\beta_j^2)$
- [B] $\min\limits_{\hat\beta}(RSS + \lambda \sum_{j=0}^{k} \hat\beta_j)$
- [C] $\min\limits_{\hat\beta}(RSS)$
- [D] пошагового исключения факторов
- [E] $\min\limits_{\hat\beta}(RSS + \lambda \sum_{j=0}^{k} |\hat\beta_j|)$
- [F] *Нет верного ответа.*

Вопрос 8 ♣
Если функция плотности удовлетворяет условиям регулярности, оценки метода максимального правдоподобия

- [A] Всегда состоятельные
- [B] Несмещённые
- [C] Могут иметь произвольное асимптотическое распределение
- [D] Всегда неотрицательны
- [E] Всегда имеют нормальное распределение
- [F] *Нет верного ответа.*

Вопрос 9 ♣
Если стандартные отклонения случайных возмущений в регрессии $Y = \beta_0+\beta_1 X_1+\ldots+\beta_k X_k+\varepsilon$ пропорциональны регрессору Z, то гетероскедастичность можно устранить

- $\boxed{\text{A}}$ Поделив все факторы в исходном уравнении на Z
- $\boxed{\text{B}}$ Поделив все факторы в исходном уравнении на $\sqrt{Z}$
- $\boxed{\text{C}}$ Умножив все факторы в исходном уравнении на $\sqrt{Z}$
- $\boxed{\text{D}}$ Умножив все факторы в исходном уравнении на Z
- $\boxed{\text{E}}$ Поделив все факторы в исходном уравнении, кроме единичного столбца, на Z
- $\boxed{\text{F}}$ *Нет верного ответа.*

Вопрос 10 ♣
Модель $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ оценивается по 500 наблюдениям, $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_{500})$. При проведении теста Уайта во вспомогательную регрессию входят исходные переменные, их квадраты и кросс-произведения. Дисперсия статистики Уайта при выполненной H_0 равна

- $\boxed{\text{A}}$ 10
- $\boxed{\text{B}}$ невозможно вычислить по имеющимся данным
- $\boxed{\text{C}}$ 18
- $\boxed{\text{D}}$ 20
- $\boxed{\text{E}}$ 9
- $\boxed{\text{F}}$ *Нет верного ответа.*

Имя, фамилия и номер группы:

. . . .

Таблица заполняется проверяющим работу!

Тест	Задача 1	Задача 2	Задача 3	Задача 4	

Отметьте верный ответ в каждом вопросе ниже:

Вопрос 1 : A B C D E F

Вопрос 2 : A B C D E F

Вопрос 3 : A B C D E F

Вопрос 4 : A B C D E F

Вопрос 5 : A B C D E F

Вопрос 6 : A B C D E F

Вопрос 7 : A B C D E F

Вопрос 8 : A B C D E F

Вопрос 9 : A B C D E F

Вопрос 10 : A B C D E F

Часть 2. Задачи.

1. На основании наблюдений получена МНК оценка уравнения регрессии $\hat{Y}_i = 0.2Z_i + 0.3W_i$ и оценка дисперсии ошибок $\hat{\sigma}^2 = 0.04$. Матрица наблюдений регрессоров имеет вид

$$X^T = \begin{pmatrix} 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 5 & 6 \end{pmatrix}.$$

   Ошибки имеют нормальное распределение.

   Постройте 95% предиктивный интервал (доверительный интервал для индивидуального прогноза) в точке $Z = -1, W = 3$.

2. В модели множественной регрессии $Y = X\beta + \varepsilon$ выполнены все предпосылки классической линейной модели кроме предпосылки о гомоскедастичности. Вектор ошибок имеет нормальное распределение, а возможная гетероскедастичность имеет вид

$$\mathrm{Var}(\varepsilon_i) = \begin{cases} \sigma_1^2, \text{ при } i \leq m; \\ \sigma_2^2, \text{ при } i > m. \end{cases}$$

   Матрица X имеет размер n на $k+1$.

   Выведите формулу статистики LR-теста для проверки гипотезы о гомоскедастичности.

3. Рассмотрим модель $Y_i = \beta X_i + \varepsilon_i$, где ε_i — независимые случайные величины с $\mathbb{E}(\varepsilon_i) = 0$ и $\mathrm{Var}(\varepsilon_i) = 2018i$.

   Найдите наиболее эффективную оценку для параметра β в классе всех линейных по Y несмещённых оценок.

4. По 1000 наблюдений Винни-Пух оценил логистическую модель $\mathbb{P}(Y_i = 1) = F(\beta_0 + \beta_1 X_i)$, где X_i — количество времени в часах, проведённое в гостях, а Y_i — факт застревания при выходе.

   Оценки параметров равны $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 2$, с оценкой ковариационной матрицы

$$\begin{pmatrix} 0.25 & 0.1 \\ 0.1 & 0.36 \end{pmatrix}.$$

   а) Проверьте значимость отдельных коэффициентов при уровне значимости 5%;
   б) Найдите предельный эффект времени, проведённого в гостях, на вероятность застрять при выходе для получасового визита;
   в) Найдите максимально возможный предельный эффект.